RAINFALL PROBABILITIES DURING THE CROP SEASON IN SOUTHERN LOWER MICHIGAN

A. H. EICHMEIER

U.S. Weather Bureau, East Lansing, Mich.

and

W. D. BATEN

Michigan State University, East Lansing, Mich.

[Manuscript received January 22, 1962; revised April 9, 1962]

ABSTRACT

Daily precipitation records for 91 years are analyzed to determine the probability of rain tomorrow (1) regardless of whether or not it is raining today, (2) given the condition that it is raining today, and (3) given the condition that it is not raining today. Considerable short-period variation was found along with a seasonal trend. A curve for seasonal trend was obtained by the least squares method. Daily variations from this curve tested by the Chi-Square method indicated that these short-term variations were not real, but due to insufficient size of sample. Also the number of n consecutive dates with probability values above or below the smoothed curve were found through Chi-Square tests to conform in frequency to the formula 0.5^n .

1. INTRODUCTION

In this paper the climatological probability of precipitation tomorrow for south central Michigan is determined for days in the crop season under three different conditions: (1) rainfall occurrence today unknown; (2) rain today; and (3) no rain today.

The criterion of 0.01 inch was used for tabulating the "rain" or "no rain" days after consultation with the Farm Crops Department at Michigan State University. Michigan's problem is more often one of having dry enough weather during haymaking and harvesting than of having sufficient moisture to break a drought and it was desired that any measurable amount be considered a day with rain. This also agrees with the U.S. Weather Bureau definition of a day with rain, so that the probabilities obtained can be used by forecasters in determining the climatological chance of having rain the next day.

2. BASIC DATA

Weather records were begun on the campus of Michigan State University in 1864 with R. C. Kedzie, professor of chemistry, taking the observations. Professor Kedzie continued to serve as observer for 38 years, making accurate and consistent readings. After his death in 1902, various members of the chemistry staff took the observations until 1910, when the Weather Bureau established an office on the campus and took over the observations. Records for some of the earlier years, namely 1867, 1875, 1892, and 1893, are missing.

Daily rainfall data for 91 years were tabulated as shown in table 1 based on rain tomorrow whether or not it was

raining today. The solid curve in figure 1 was constructed using the probability data (actual values) from table 1 showing many short term variations as well as a seasonal trend. The dashed smooth curve in figure 1 was determined by the least squares method, using the formula

$$y=a+bx+cx^2+dx^3$$
 (1)

where y equals the percentage of days with precipitation for a particular date and x equals the number of days from (July 1). Values found for a, b, c, and d are as follows: a=30.793808, b=-0.147624, c=0.000354, d=0.000019 The smoothed values of probability given by equation (1) and used in constructing the curve are given in table 1 (smoothed values).

3. SIGNIFICANCE TESTS

A test of significance was now in order to determine whether the day-to-day variations from the smoothed curve were real or due to the restricted size of the sample. The observed data and the expected data (actual frequencies, not percentages) were used to make daily Chi-Square tests [1]. Then from an addition theorem for Chi-Square, the sum of the 183 Chi-Squares was found to be 194.36, which is non-significant, with 183 degrees of freedom. This indicates that the samples were drawn from a homogeneous population and that variation between the observed data and the constructed curve are probably due to sampling variations and not to population differences.

The limiting value at the 5 percent level can be found (approximately) by the formula, $n+0.85+1.645\sqrt{2n-1}$; placing n=183 this value is 215.66.

Table 1.—Precipitation frequency based on 91 years of record, East Lansing, Mich.

	TABLE 1. Treespitation frequency bused on 51 years of record, Base Banesing, Mich.																						
April	No. times rain tomorrow Actual frequency	Prob. (%) rain tomorrow Actual values	Prob. (%) rain tomorrow Smoothed values	May	No. times rain tomorrow Actual frequency	Prob. (%) rain tomorrow Actual values	Prob. (%) rain tomorrow Smoothed values	June	No. times rain tomorrow Actual frequency	Prob. (%) rain tomorrow Actual values	Prob. (%) rain tomorrow Smoothed values	July	No. times rain tomorrow Actual frequency	Prob. (%) rain tomorrow Actual values	Prob. (%) rain tomorrow Smoothed values	August	No. times rain tomorrow Actual frequency	Prob. (%) rain tomorrow Actual values	Prob. (%) rain tomorrow Smoothed values	September	No. times rain tomorrow Actual frequency	Prob. (%) rain tomorrow Actual values	Prob. (%) rain tomorrow Smoothed values
1 22 3 4 5 6 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 22 23 24 25 26 27 28 29 30 30 30 30 30 30 30 30 30 30 30 30 30	31 30 30 36 34 36 26 25 31 39 30 32 29 35 39 31 31 33 32 24 42 28 39 34 29 28	34 33 33 340 57 40 29 27 34 43 33 33 35 32 38 43 36 35 26 31 46 31 32 32 32 33 31 31 31 31 31 31 31 31 31 31 31 31	33. 0 33. 2 33. 5 34. 0 34. 4 34. 8 34. 9 35. 0 35. 2 35. 5 35. 7 35. 5 35. 7 36. 2 36. 2 36. 2 36. 3 36. 4 36. 6 36. 6 36. 6 36. 8 36. 8	1 2 2 3 4 4 5 6 6 7 7 8 8 9 9 100 111 122 133 144 155 6 17 18 19 20 22 23 32 24 25 25 29 30 31 31	33 33 27 24 37 29 38 38 38 39 31 29 37 40 39 32 35 41 31 34 43 43 43 25 40 39 25 40 39 39 31 40 31 40 40 40 40 40 40 40 40 40 40 40 40 40	36 36 36 30 32 42 43 34 34 34 32 41 44 43 35 38 45 37 47 48 33 32 43 37 37	36. 9 36. 9 36. 9 36. 9 36. 9 36. 8 36. 8 36. 8 36. 7 36. 7 36. 7 36. 5 36. 4 36. 3 36. 2 36. 2 36. 1 36. 0 35. 9 35. 6 35. 6 35. 6 35. 6 35. 6 35. 6	1 2 2 3 4 4 5 5 6 7 7 8 9 100 111 122 13 14 15 6 17 18 19 20 12 22 23 24 25 26 27 28 29 30	30 27 31 28 40 33 36 23 31 31 25 25 37 28 36 31 23 31 22 29 34 42 27 27 27 22 32 22 32 23 23 23 23 23 23 24 23 24 24 25 26 27 27 27 27 27 27 27 27 27 27 27 27 27	33 30 34 34 36 40 25 34 34 31 40 34 33 34 25 37 30 30 32 31 31 31 31 31 31 31 31 31 31 31 31 31	35. 0 34. 9 34. 8 34. 7 34. 6 34. 3 34. 1 33. 8 33. 8 33. 8 33. 5 33. 5 32. 9 32. 5 32. 4 32. 1 31. 8 31. 7 31. 8 31. 7 31. 8 31. 7 31. 8 31. 7 31. 8 31. 9	1 2 3 3 4 4 5 5 6 7 7 8 9 9 100 111 122 133 14 15 16 17 18 19 20 12 22 23 30 30 31	300 229 229 177 228 31 224 222 226 229 225 221 225 23 23 24 25 229 225 24 226 229 227 227 228 229 229 229 221 220 220 221 221 222 223 233 244 247 257 267 277 277 277 277 277 277 277 277 27	33 52 32 31 31 29 34 31 26 24 29 32 21 27 25 21 23 29 29 29 29 29 29 21 21 23 29 21 21 23 24 25 31 27 25 31 27 28 29 21 21 21 22 23 24 25 21 26 27 27 28 29 29 29 29 20 20 20 20 20 20 20 20 20 20	30. 8 30. 6 30. 5 30. 4 30. 3 30. 1 30. 0 29. 8 29. 7 29. 5 29. 2 29. 1 29. 0 28. 9 28. 7 28. 6 28. 5 28. 4 28. 2 27. 8 27. 8 27. 8 27. 8 27. 8 27. 4 27. 4 27. 4 27. 3 27. 2	1 2 2 3 3 4 4 5 5 6 7 7 8 8 9 9 100 111 122 133 144 15 6 17 7 18 19 20 12 22 23 30 24 22 5 6 27 28 8 29 30 31	33 27 29 26 26 28 23 27 23 21 25 21 24 24 24 22 26 33 21 22 20 20 21 22 23 24 24 24 22 26 23 26 23 26 23 27 27 27 27 27 27 27 27 27 27 27 27 27	36 30 32 29 29 31 25 25 25 30 25 26 26 26 24 29 36 23 24 22 23 22 24 25 29 29 29 20 20 20 20 20 20 20 20 20 20 20 20 20	27. 1 27. 0 26. 9 26. 9 26. 8 26. 8 26. 7 26. 6 26. 7 26. 8 26. 9 27. 0 27. 1 27. 1 27. 1 27. 2 27. 4	1 2 2 3 3 4 4 5 5 6 7 7 8 8 9 9 100 111 112 13 13 14 15 6 17 7 18 19 20 1 22 23 24 22 5 22 23 30	27 19 27 30 30 26 23 34 22 28 33 34 24 24 24 30 28 31 27 27 27 24 35 31 27 27 27 27 21 21 21 21 21 21 21 21 21 21 21 21 21	30 211 303 333 29 255 37 38 34 35 31 36 26 33 31 25 30 26 37 38 39 26 31 31 32 32 32 32 32 32 32 32 32 32 32 32 32	27. 5 27. 6 27. 8 28. 0 28. 0 28. 1 28. 3 28. 4 28. 6 29. 2 29. 4 29. 2 29. 4 30. 1 30. 4 30. 1 31. 1 31. 7 32. 3 32. 7 33. 3 34. 4 34. 4

A second test was made to determine whether or not the number of consecutive dates with sample values above the average or below the average as shown by the smooth curve differed in frequency from those which would be expected by sampling variation. If it is assumed that the true population is correctly indicated by the smooth curve, and that approximately one-half of the sample values are above and one-half below the curve values, then 0.5^n will give the probability of n consecutive dates above or n consecutive dates below the average. Table 2 shows the frequency of occurrence, actual and theoretical (expected), of the consecutive dates with values above, or below, the computed (smoothed) values. This table and figure 2 show that these two sets of values agree closely. However, a test of significance was set up using Chi-Square as defined by:

$$\chi^2 = \sum_{i=1}^{m} \left\{ (O_i - H_i)^2 / H_i \right\}$$
 (2)

The number of occurrences in each category is O_1 , O_2 , O_3 , . . . , O_m . The null hypothesis is made that this sample belongs to a population in which the number of expected occurrences is H_1 , H_2 , H_3 , . . . , H_m .

The last six categories (table 2) are grouped to increase the size of H_5 . Also, since the H_i are small, the values of the absolute discrepancies are reduced by applying the Yates correction of 0.5. Then equation (2) becomes:

$$\chi^2 = \sum_{i=1}^{5} \left\{ (|O_i - H_i| - 0.5)^2 / H_i \right\} \tag{3}$$

With data from table 2, equation (3) gives $\chi^2 = 1.74$.

The only restriction is the condition that the total number of individuals in the hypothetical distribution has to be the same as that in the sample distribution. In the above instance there are five categories and the only restriction is the given total. The number of cases in the first four categories of the hypothetical distribution is almost unrestricted, and only the number in the last group is fixed by the given total [2]. Therefore, there are four degrees of freedom. The limiting value of Chi-Square at the 5 percent level with four degrees of freedom is 9.49. The value of 1.74 is considerably less than this;

Table 2.—Actual values consecutively above or consecutively below computed values. 91 years of record, East Lansiny, Mich.

Number of consecu- tive days	Actual frequency	Expected frequency			
1	36	42. 5			
2	26	21, 2			
3	11	10. €			
4	6	5. 3			
5	2	2. 7			
6	2	1, 3			
7	2	0.7			
3	0	0.3			
9	. 0	0, 2			
10	. 0	0. 1			

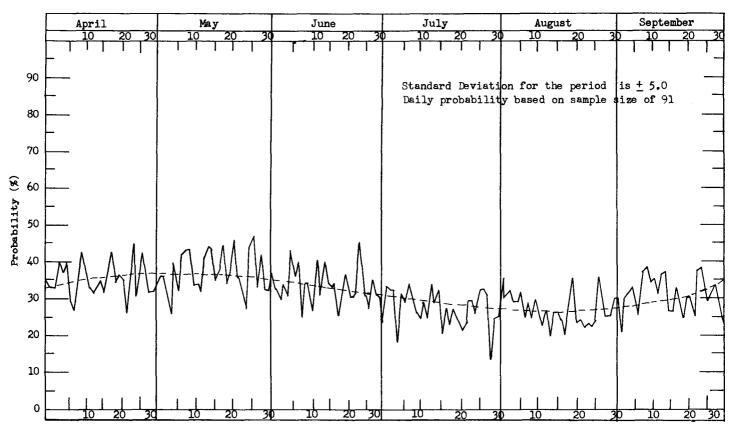


Figure 1.—Probability of rain tomorrow based on condition of rain or no rain today. 91 years of record, East Lansing, Mich. Standard deviation for the period is ± 5.0 . Daily probability based on sample size of 91.

hence the null hypothesis is accepted. This indicates also, that there is an absence of well-defined precipitation singularities during the crop season in south central lower Michigan.

4. CONDITIONAL PROBABILITIES

The probabilities thus far were based on all dates regardless of precipitation or no precipitation on the starting day and can be used to give climatological expectancies on a long-range basis. However, on a short-range basis and as a supplementary aid to the forecaster, two conditional probabilities were next developed: based on the conditions (1) rain today, and (2) no rain today, what is the probability of rain tomorrow? As in table 1, data used in developing these conditional probabilities are independent since frequencies are computed (under the condition imposed) for occurrence on the same date in different years and not for consecutive days in any one year.

Since the frequency of occurrence of these happenings varied from date to date the size of the sample also varied and in figures 3 and 4 it was necessary to change frequencies to percentages before developing the curve. Percentages were also used in developing the curve for figure 1, but frequencies could have been used there.

Daily Chi-Square values were obtained from the actual and expected frequencies (not percentages). Sums of Chi-Square based on data from figures 3 and 4 are 192.67 and 167.27 respectively, again below the limiting value of 215.66 for the 5 percent level, and therefore also indi-

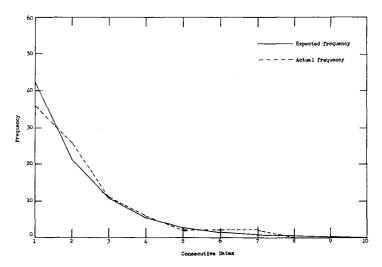


Figure 2.—Frequency of occurrence of consecutive dates above the computed value or below the computed value in figure 1. 91 years of record, East Lansing, Mich.

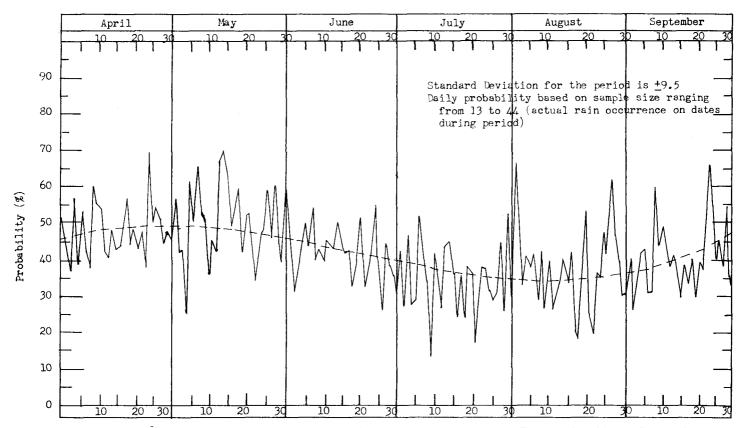


FIGURE 3.—Probability of rain tomorrow based on condition of rain today. 91 years of record, East Lansing, Mich. Standard deviation for the period is ± 9.5 . Daily probability based on sample size ranging from 13 to 44 (actual rain occurrences on dates during period).

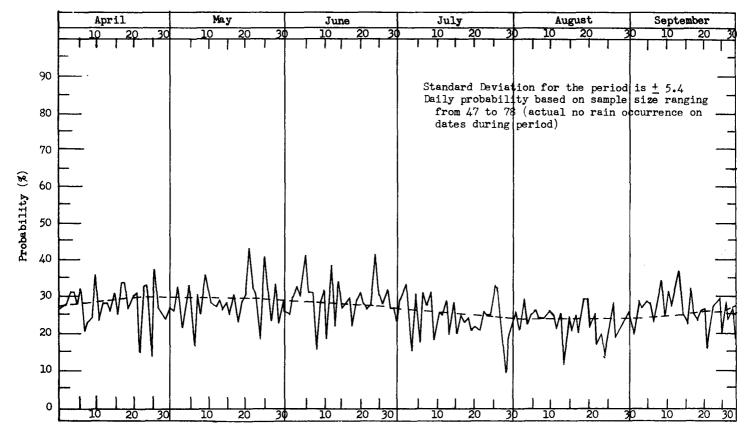


FIGURE 4.—Probability of rain tomorrow based on condition of no rain today. 91 years of record, East Lansing, Mich. Standard deviation for the period is ±5.4. Daily probability based on sample size ranging from 47 to 78 (actual no rain occurrences on dates during period).

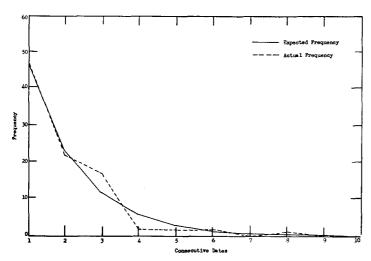


Figure 5.—Consecutive dates above the computed value or below the computed value in figure 3.—91 years of record, East Lansing, Mich

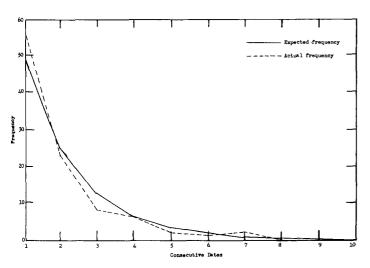


Figure 6.—Consecutive dates above the computed value or below the computed value in figure 4. 91 years of record, East Lansing, Mich.

cate that the samples were drawn from homogeneous populations with true values estimated by the constructed curves. Figures 5 and 6 show the actual and expected frequencies of number of consecutive dates above or below the computed values. Chi-Square tests of significance by setting up the null hypothesis (as explained earlier with data from table 2) again show no significant difference between the observed and hypothetical distributions.

Table 3.—Computed probability of rain tomorrow based on: All cases; rain today; and no rain today (average for stated periods). 91 years of record, East Lansing, Mich.

Period	Prob. (%) rain tomorrow, all cases	Prob. (%) rain tomorrow, given the condition of rain today	Prob. (%) rain tomerrow, given the condition of no rain today	Period	Prob. (%) rain tomorrow, all cases	Prob. (%) rain tomorrow, given the condition of rain today	Prob. (%) rain tomorrow, given the condition of no rain today		
April 1-10	34. 0 35. 7 36. 6 36. 9 36. 5 35. 7 34. 5 33. 1 31. 6	47. 0 48. 5 49. 1 49. 0 48. 1 46. 6 44. 7 42. 7 40. 6	27. 9 28. 9 29. 5 29. 6 29. 5 29. 0 28. 4 27. 7 26. 9	July 1-10 July 11-20 July 21-31 August 1-10 August 11-20 August 21-31 September 1-10 September 21-31 September 21-30.	30. 2 28. 9 27. 7 26. 9 26. 6 27. 0 28. 1 30. 0 32. 9	38. 6 36. 8 35. 4 34. 6 34. 5 35. 4 37. 4 40. 5	26. 2 25. 4 24. 7 24. 3 24. 0 24. 2 24. 6 25. 5 26. 9		

5. SUMMARIZATION

Table 3 summarizes the results and gives the probability of rain tomorrow under the three following conditions: (1) when the rainfall occurrence today is unknown, (2) when the condition of rain today is known, (3) when the condition of no rain today is known. For practical purposes 10-day averages give a close approximation of the probabilities involved and table 3 gives such averages for three periods in each month. For example, on May 6, should there be no rain, the table shows rain could be expected on the following day about 3 out of 10 years, but should there be rain on May 6, the chance of rain on the following day is about 50-50.

6. CONCLUSIONS

Probabilities based on a computed curve in which seasonal variation is recognized, but in which day-to-day variation is eliminated, are believed very near the true population mean when based on samples of size used in this paper. Short period precipitation singularities are not indicated during the crop season in south central lower Michigan. The probability of having rain tomorrow, given the condition that it is raining today, is considerably greater than if it is not raining today.

REFERENCES

- George W. Snedecor, Statistical Methods, The Iowa State College Press, Ames, Iowa, 1956.
- Hans A. Panofsky and Glenn W. Brier, Some Applications of Statistics to Meteorology, The Pennsylvania State University, University Park, Pennsylvania, 1958.